



Bases lexicales multilingues : traitement des acronymes

Ying Zhang, Mathieu Mangeot

► To cite this version:

Ying Zhang, Mathieu Mangeot. Bases lexicales multilingues : traitement des acronymes. première journée " Jeunes Chercheurs " du réseau Lexicologie Terminologie Traduction (JCLTT 2013), 2013, Bruxelles, Belgique. pp.6. hal-00953763

HAL Id: hal-00953763

<https://hal.science/hal-00953763>

Submitted on 1 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bases lexicales multilingues : traitement des acronymes

Ying ZHANG^{1*} et Mathieu MANGEOT^{2*}

*GETALP-LIG ¹Université Joseph Fourier ²Université de Savoie

Résumé

La gestion des terminologies pose encore des problèmes, en particulier pour la gestion des situations compliquées comme les acronymes. Dans cet article, nous proposons une solution en reliant à un seul référent plusieurs termes différents via les notions de pivot et de prolexème. Ces notions permettent par exemple de faire le lien entre plusieurs termes qui désignent un même et unique référent : Nations Unies, ONU, Organisation des Nations Unies et onusien. Il existe, une plate-forme générique de gestion de bases lexicales, Jibiki, permettant de gérer n'importe quel type de structure (macro et microstructure). Nous avons implémenté une nouvelle macrostructure de ProAxie dans la plate-forme Jibiki pour réaliser les gestions des acronymes.

Mots-clés :

base lexicale multilingue, macrostructure, Jibiki, acronyme, Common Dictionary Markup, entrepôt de données linguistiques

1 Introduction

Cet article concerne la gestion de terminologies multilingues. Le problème abordé dans cet article est celui de l'association de plusieurs termes d'une même langue à un même référent. Par exemple un terme et son acronyme désignent le même référent. Dans un contexte multilingue, la difficulté est d'établir une correspondance entre ces termes. Il peut par exemple y avoir dans une même langue plusieurs acronymes pour un même terme, voire aucun. Quelles solutions mettre en place de façon à choisir, pour un terme donné dans une langue donnée, le meilleur équivalent dans une langue cible ? Cette recherche est motivée par un besoin réel d'une entreprise dans la gestion de sa terminologie multilingue.

L'intérêt de cette recherche est de définir un cadre théorique composé d'une nouvelle macrostructure basée sur des concepts existants et sur la définition de nouveaux concepts. Ce cadre sera validé ensuite par une expérimentation pratique à l'aide d'un outil générique de gestion de bases lexicales.

Cet article est organisé de la façon suivante. Dans la section 2, nous présentons les problèmes et les difficultés de gestion des terminologies riches. Dans la section 3, nous présentons

les macrostructures préconisées pour les données. La section 4 présente les outils utilisés et l'implémentation de la macrostructure. La section 5 présente les résultats de l'implémentation et de l'utilisation. Enfin, nous concluons et donnons quelques perspectives pour la gestion des terminologies riches.

2 Problématique

Les liens entre les termes sont compliqués. Il existe la possibilité de lier plusieurs termes différents à un seul référent : « Jean-Paul II » et « Karol Jozef Wojtyla » en français, ou en anglais « John Paul II » et « Karol Jozef Wojtyla ». De même, certains liens évoluent avec le temps : le pape désignait « Jean-Paul II » en 2004 et « Benoît XVI » en 2012. Des pays parlant la même langue (p. ex : France et Suisse romande) peuvent également utiliser des mots différents pour le même concept. Par exemple, « chien renifleur » et « chien drogue ». Inversement, le même terme peut désigner des concepts différents : Dans la province de langue allemande de Bolzano en Italie, le « Landeshauptmann » est le président du conseil provincial, avec des compétences beaucoup plus limitées que le « Landeshauptmann » autrichien, qui est à la tête de l'un des États (Land) de la fédération autrichienne.

Pour la gestion des acronymes, les liens riches sont plus répandus et plus complexes. Dans l'article introduisant la notion de Prolexème [MT00], il a été discuté qu'il y a des lexies ayant des acronymes dans certaines langues, mais pas dans d'autres. Par exemple, en français, il y a « organisation des nations unies »¹, on peut aussi dire « Nations unies », « ONU » ou « onusien ». En anglais, on a « United Nations » et son acronyme « UN ». En chinois, on a « 联合国 » qui est la seule lexie pour ce sens, et il n'y a pas d'acronyme. Quand on construit des ressources lexicales, comment créer les liens entre les acronymes et les mots de définition ?

Dans les chapitres suivants, nous parlerons des idées pour résoudre les problèmes décrits ci-dessus.

3 Données : choix de la macrostructure

Lors de toute discussion scientifique, il est primordial de bien s'entendre sur les termes utilisés. C'est pourquoi nous commencerons par définir les termes et concepts principaux que nous utiliserons par la suite.

Un dictionnaire est composé d'un ou plusieurs volumes reliés entre eux par des liens qui sont le plus souvent des liens de traduction. Un volume est un ensemble d'articles comportant des mots-

¹ Nous avons repris exactement la terminologie française de Prolexbase et les concepts.

vedettes de la même langue. Un article comporte au moins un mot-vedette et le plus souvent d'autres informations (prononciation, classe grammaticale, définition, exemples, etc.). La structure des articles est appelée microstructure. L'organisation des volumes qui compose la structure d'un dictionnaire est appelée macrostructure. La macrostructure la plus simple est celle d'un dictionnaire monolingue ne comportant qu'un seul volume. Pour les dictionnaires bilingues langue A \leftrightarrow langue B, on peut trouver des macrostructures avec deux volumes : un volume langue A \rightarrow langue B et un volume miroir langue B \rightarrow langue A. Ces macrostructures constituent l'essentiel des dictionnaires imprimés. L'avènement de l'outil électronique permet de s'abstraire des contraintes liées à l'impression, notamment la représentation restreinte à deux dimensions. Il est alors possible de concevoir des macrostructures plus complexes utilisant par exemple des volumes pivot. Le dictionnaire devient alors une base lexicale à plusieurs dimensions d'où il est ensuite possible d'extraire des vues spécifiques permettant de retrouver le format initial des dictionnaires imprimés.

Nous détaillerons trois macrostructures de ce type par la suite.

3.1 Macrostructure pivot

Le projet Papillon lancé en 2000 [MTMMEP00], a pour but de construire une ressource lexicale pour plusieurs langues dont au moins l'anglais, le français et le japonais. Les macrostructures bilingues traditionnelles obligeant à construire un dictionnaire par couple de langues, le nombre de dictionnaires croît de manière factorielle par rapport au nombre de langues en présence. Cette solution devient rapidement ingérable. Il fallait donc en trouver une nouvelle, un dictionnaire multilingue à structure pivot : un volume monolingue pour chaque langue et un volume pivot (ou volume interlingue) au centre regroupant les liens entre les articles [GSMM01a]. La microstructure des articles monolingues reprend le concept de lexie défini dans la lexicographie explicative et combinatoire [MLCAPA00] issue de la théorie sens-texte. Chaque article décrit une lexie. Une lexie est une unité lexicale (ou sens de mot) qui est représentée soit par un lexème (regroupement de mots-forme), soit par une locution nominale.

Chaque lexie est reliée par un lien interlingue à une axie (ou acception interlingue). Les axes sont contenues dans le volume pivot. Chaque axie regroupe les équivalents dans plusieurs langues d'une même lexie (ou sens de mot).

Les concepts d'axie et de structure pivot ont été définis pour le projet Papillon et ensuite repris dans la norme Lexical Markup Framework [Francopoulo09].

3.2 Macrostructure Pivax

La macrostructure pivot vue précédemment permet de résoudre le problème posé par une

situation multilingue. Au niveau monolingue, par contre, on doit se contenter d'un seul volume pour chaque langue, ce qui constitue une limitation importante. Dans sa thèse, Hong-Thai Nguyen [ThesisHTN] avait le projet de construire une base lexicale avec des ressources issues de systèmes de traduction. Chaque système de traduction utilise un format qui lui est propre pour ses dictionnaires lui permettant de stocker des variables spécifiques. Si l'on veut regrouper ces dictionnaires, il n'est pas souhaitable de fusionner tous les dictionnaires de chaque langue. On peut par contre regrouper les lexies identiques d'une même langue dans un même objet. C'est le rôle des axèmes, ou acceptions monolingues.

Pour chaque langue ou espace lexical, on aura alors un volume pour chaque système de traduction présent (Systran, Ariane, etc.) et un seul volume d'axèmes. Chaque lexie de chaque volume est reliée à un axème. Chaque axème est à son tour relié à une axie dans le volume pivot central.

Cette structure ainsi constituée est une structure pivax ou structure pivot à étages [MMHTN09]. Elle peut être utilisée également pour gérer différentes versions d'un même volume dans un espace lexical.

3.3 Macrostructure ProAxie

Cette macrostructure a pour but de résoudre le problème de faire le lien entre plusieurs termes qui désignent un même et unique référent. Pour les gestions des acronymes, les liens riches sont plus répandus et plus complexes. Dans l'article introduisant la notion de Prolexème, il a été discuté qu'il y a des lexies ayant des acronymes dans certaines langues, mais pas dans d'autres. Donc, la macrostructure de la gestion des acronymes devra nous permettre de

- gérer les liens plus complexes et
- gérer les multiples volumes dans la même langue.

Pour implémenter la gestion des acronymes, nous avons proposé une nouvelle macrostructure avec deux notions: proaxie et prolexème [MT00]. Nous avons également besoin des concepts d'axie [MTMMEP00] et de lexie [MLCAPA00].

3.3.1 Prolexème

Il y a un seul volume de prolexèmes pour chaque langue. Dans ce volume, les prolexèmes regroupent les lexies. Les liens bidirectionnels entre les lexies et leurs prolexèmes sont marqués avec une étiquette (alias, acronyme, définition, etc.). Un sens lexical est une entrée de prolexème qui lie les différentes lexies de ce même sens. Par exemple, l'entrée de prolexème est

« fr.organisation_des_nations_unies.1 » qui lie l'entrée des lexies :

- « ONU » avec une étiquette « acronyme »
- « nations unies » avec une étiquette « alias »
- « onusien » avec une étiquette « dérivée »
- « organisation des nations unies » avec une étiquette « définition ».

3.3.2 Proaxie

Il y a un seul volume de proaxies dans un dictionnaire. Les proaxies regroupent les prolexèmes de langues différentes dans un même sens. Par exemple, dans un dictionnaire trilingue : français, anglais et chinois. L'entrée de proaxie « proaxie.united_nations.1 » relie

- l'entrée « fra.organisation_des_nations_unies.1 » du volume des prolexèmes français,
- l'entrée « eng.united_nations.1 » du volume des prolexèmes anglais,
- l'entrée « zho.联合国.1 » du volume des prolexèmes chinois.

Les liens entre l'entrée de proaxie et les entrées de prolexèmes sont bidirectionnels.

3.3.3 Conception globale

Dans cette macrostructure, nous avons deux couches : une couche basique et une couche « Pro ». Dans la couche basique, nous avons deux types de volume : volume des lexies et volume des axes. Dans la couche « Pro », nous avons également deux types de volume : volume des prolexèmes et volume des proaxies.

Grâce au volume d'axes, nous pouvons faire les liens entre les lexies qui correspondent exactement, comme l'acronyme français « ONU » relié avec l'acronyme anglais « UN ». Grâce à la couche « Pro », nous pouvons proposer en traduction des lexies des langues cibles de même sens. Voir le figure 1 - « Exemple de ProAxie ».

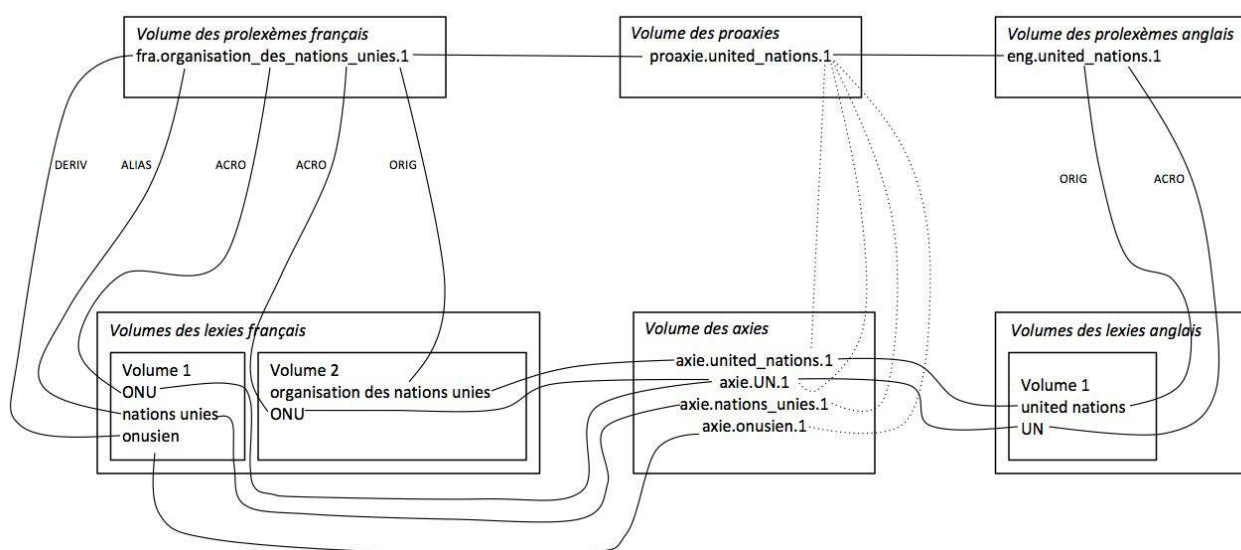


Figure 1 : Exemple de ProAxie

La notion d'étiquette a pour but de proposer les meilleures traductions. Par exemple, le japonais « 国際連合 » est la lexie de même sens que « Organisation des nations unies », son acronyme est « 国連 » comme l'indique le table 1 - « Exemple de quatre langues pour le sens « Organisation des nations unies » ». Cet acronyme utilise le premier et le troisième caractère d'idéogramme ou kanji, ce qui est différent des initiales de la lexie de définition. Il existe peut-être une langue qui a deux acronymes, l'un correspondant à l'acronyme des initiales, l'autre correspondant à une sélection de caractères ou de mots. Donc, nous avons décidé de ne pas lier ces deux types acronymes avec une même axie. Voir le figure 2 - « Liens entre les lexies et les axes ».

Langue	Anglais	Français	Chinois	Japonais
Définition	United Nations	Organisation des Nations Unies	联合国	国際連合
Acronyme	UN	ONU		国連
Alias		Nations Unies		
Dérivée		onusien		

Table 1 - Exemple en quatre langues pour le sens « Organisation des Nations Unies »

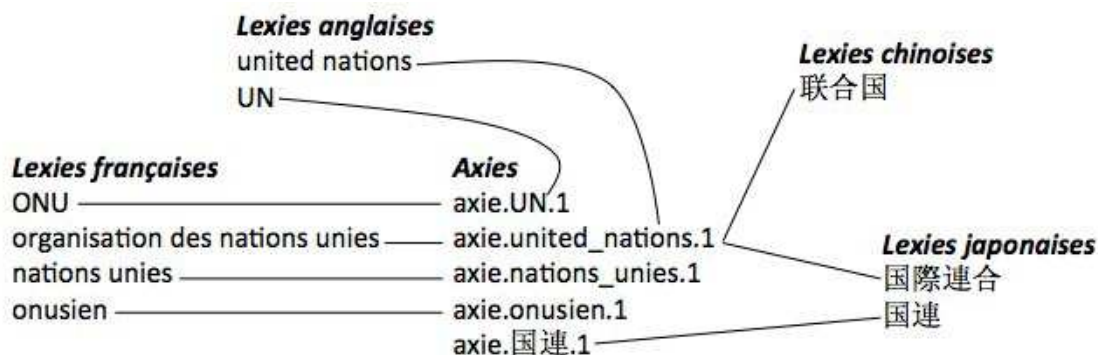


Figure 2 : Liens entre les lexies et les axes

Lorsqu'on cherche les liens de la lexie « ONU » du français vers l'anglais, vers le japonais et vers le chinois, nous proposons trois niveaux de traduction classés selon la précision obtenue :

- Vers l'anglais : « ONU »→« UN ». Le système trouve une lexie directe par le volume des axes. C'est le premier niveau de traduction et le plus précis.
- Vers le japonais : « ONU »→« 国連 ». Le système cherche le lien dans le volume des prolexèmes français avec l'étiquette « Acro ». Puis il trouve le lien dans les proaxies, ensuite il suit le lien de prolexème japonais, enfin il arrive au volume des lexies japonais, et il trouve une lexie avec l'étiquette « Acro ». Donc la lexie proposée du deuxième niveau de langue cible est cet acronyme. Le deuxième niveau de traduction comprend toujours le premier niveau de traduction. C'est-à-dire que « ONU » et « UN » ont la même étiquette « Acro », donc le lien « ONU »→« UN » correspondu également au deuxième niveau de traduction.
- Vers le chinois : « ONU »→« 联合国 ». Le système trouve les lexies par prolexème et proaxie sans étiquette. Ces lexies proposées constituent le troisième niveau, le moins précis. Le troisième niveau de traduction comprend les niveaux précédents.

La quantité de lexies de résultat est augmentée suivant les niveaux de traduction du premier vers le troisième. Nous avons également utilisé l'exemple de traduction du terme « ONU » pour l'expliquer précisément, voir la table 2 - « Trois niveaux de traduction : les traductions du terme « ONU » ».

	Premier niveau	Deuxième niveau	Troisième niveau
Anglais	UN	UN	UN, United Nations
Japonais		国連	国連, 国際連合
Chinois			联合国

Table 2 - Trois niveaux de traduction : les traductions du terme « ONU »

Dans certaines situations, une base lexicale (un dictionnaire) a plusieurs volumes pour une seule langue. Par exemple, lorsqu'il y a plusieurs versions d'édition ou que la ressource lexicale est créée par un système de Traduction Automatique, on trouvera par exemple un volume provenant de Systran, un volume de Google, un volume d'IATE, etc. Notre macrostructure permet de gérer plusieurs volumes dans une même langue. Voir la figure 3 « macrostructure de ProAxie ». Étant donnée une langue, il existe un ou plusieurs volumes de lexies, mais un seul volume de prolexèmes. Pour un dictionnaire il y a un seul volume de proaxies et un seul volume d'axies. Les entrées des lexies sont liées aux entrées de prolexèmes et sont aussi reliées aux entrées d'axies. De plus, les entrées de prolexèmes sont reliées avec les entrées de proaxies, et vice-versa.

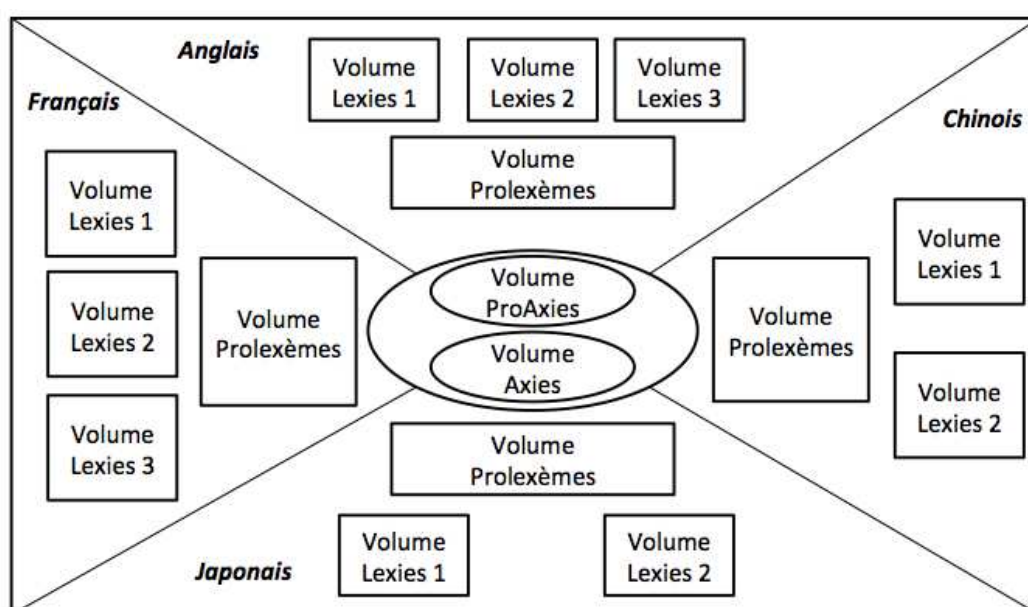


Figure 3 : Macrostructure de ProAxie

4 Outils nécessaires : plateformes de manipulation

4.1 Plateforme Jibiki version 1

Pour implémenter la macrostructure de ProAxie, nous avons utilisé la plate-forme Jibiki. Elle permet la construction de sites web de communauté dédiées à la construction de bases lexicales multilingues. Cette plate-forme a été développée principalement par Mathieu Mangeot et Gilles Sérasset [MM06]. Cette plate-forme est utilisée dans divers projets (projet LexALP, projet Papillon, projet GDEF, etc.). Le code est disponible en source ouverte et téléchargeable gratuitement sur ligforge.imag.fr par SVN. Avec cette plate-forme, on peut faire les manipulations d'import, export, édition, modification et recherche dans des bases lexicales. On peut aussi gérer les contributions.

Jibiki est une plate-forme générique, elle permet de traiter presque toutes les ressources lexicales de type XML en utilisant différentes microstructures et macrostructures. Nous avons déjà parlé de la macrostructure dans la section 3. La microstructure est la structure des entrées. On utilise des pointeurs CDM (Common Dictionary Markup) [MM02] pour gérer n'importe quel type de microstructure sans la modifier. Les pointeurs sont utilisés également pour indexer des parties d'information spécifiques et permettre ensuite une recherche multi-critères. Cette structure est stockée dans un fichier de métadonnées sous forme XML. Les fichiers de métadonnées ont pour but de faciliter l'import de ressources lexicales. Il y a un fichier de métadonnées pour chaque ressource. Dans ce fichier, sont décrites les informations sur cette ressource : les langues source et cibles, l'auteur, les noms et fichiers de volumes, etc. Pour chaque volume, il existe un fichier de métadonnées. Dans ces fichiers, sont décrites toutes les informations des volumes, y compris les pointeurs CDM. Pour chaque pointeur CDM, on indique le chemin XPath vers l'élément correspondant dans la microstructure XML. Voir la figure 4 - « Exemple de CDM ». Les liens de traduction sont à ce stade traités comme des pointeurs CDM classiques (voir la figure 4, cdm-translation-ref).

```
<cdm-elements>
  <cdm-volume xpath="/g:volume"/>
  <cdm-entry xpath="/g:volume/g:article"/>
  <cdm-entry-id xpath="/g:volume/g:article/@g:id" />
  <cdm-headword xpath="/g:volume/g:article/g:vedette/g:mot/text()" d:lang="fra" />
  <cdm-pronunciation xpath="//g:prononciation/text()" d:lang="fra" />
  <cdm-pos xpath="//g:cat-gram/text()" d:lang="fra" />
  <cdm-definition xpath="/g:volume/g:article/g:vedette/g:mot/text()"/>
  <cdm-translation-ref xpath="//g:mot-princ/g:value/text()" d:lang="fra"/>
</cdm-elements>
```

Figure 4 : Exemple de CDM

La version 1 de Jibiki présentait plusieurs limitations :

- Les fichiers de métadonnées utilisent des chemins Xpath pour décrire les fichier XML. La création des métadonnées peut être difficile, en particulier pour un non-informaticien.
- Les liens de traduction sont traités avec des pointeurs CDM, comme des éléments d'information classiques. Ces liens sont simples. Il n'y a pas de possibilité de décrire des liens entre plusieurs volumes différents. Il n'est pas non plus possible d'ajouter des attributs (poids, étiquette, type, volume cible etc.) sur les liens.
- Jibiki est utilisé pour plusieurs macrostructures. Pour chaque macrostructure, il est nécessaire de re-coder une partie du programme pour réaliser les différents types de

liens.

4.2 Gestion des données et métadonnées : iPoLex

Une fois la première version de Jibiki réalisée, nous nous sommes rendus compte d'un autre problème soulevé par les ressources : que faire des fichiers de données brutes au format textuel avant importation et après exportation de la plate-forme ? De plus, la rédaction de fichiers XML de métadonnées décrivant chaque ressource est très délicate et réservée à des informaticiens avertis. La gestion des différentes versions d'une même ressource pose également problème. Enfin, les ressources au format textuel peuvent être utilisées directement par d'autres applications. Il convient donc d'organiser leur stockage dans de bonnes conditions.

L'idée s'est alors imposée de concevoir un entrepôt de données lexicales permettant de stocker les données brutes au format textuel ainsi que de définir les métadonnées d'une nouvelle ressource à l'aide d'une interface. Ce projet, nommé iPoLex avait comme objectif d'être simple et rustique.

L'entrepôt est accessible par le Web. Les interfaces sont programmées en PHP sans connexion à une base de données. Toutes les informations sont stockées dans des fichiers texte. La page d'accueil affiche la liste des ressources disponibles et permet d'effectuer différents tris (format, langue source, langues cibles, etc).

L'ajout d'une nouvelle ressource se fait en trois étapes : d'abord la description de la ressource avec ses métadonnées (langues source et cibles, domaine, auteur, date de création, format, taille des fichiers texte, etc.) puis, pour chaque volume de la ressource, la description de ses métadonnées (langue source, nombre de mots-vedettes, version, etc.) ainsi que celle des pointeurs CDM permettant d'indexer le volume après importation dans la base Jibiki. La description des ressources se termine par la création d'un répertoire sur le serveur et la génération de fichiers de méta-données.

La dernière étape est celle du téléversement des fichiers de données sur le serveur. Pour cela, l'entrepôt peut se monter comme répertoire distant grâce au protocole WebDAV. Une fois les fichiers de données téléversés sur le serveur, l'import dans une plate-forme Jibiki se fait très facilement en indiquant dans l'interface de Jibiki l'URL du fichier de métadonnées de la ressource à importer.

Nous prévoyons d'étendre ce concept d'entrepôt de données à celui d'entrepôt de services. En effet, pour la plupart des ressources stockées dans l'entrepôt, il existe des services associés comme par exemple des moulinettes de conversion d'un format à un autre. Ces scripts sont en général prévus pour une version particulière des données. S'ils ne sont pas documentés précisément et

stockés avec les données, il est très difficile, voire impossible de réexécuter les processus divers associés aux données.

4.3 Gestion des liens riches : extension Jibiki-Pivax

La gestion des liens riches correspond aux liens avec des attributs, comme le volume cible, le poids, le type (axi, final), la langue, une étiquette libre, etc. Pour réaliser l'implémentation de liens riches, nous avons séparé le module de traitement des liens de celui des autres pointeurs CDM.

La réalisation de programme est basée sur deux algorithmes. Le premier est l'algorithme de collecte des liens, le deuxième est l'algorithme de construction du résultat. Le premier algorithme permet de chercher tous les liens possibles dans l'ensemble des liens riches de tous les volumes pour une entrée recherchée. Le deuxième algorithme nous permet de réaliser le parcours des liens riches. Il s'agit des étapes suivantes :

- chercher le lien vers les axes puis vers les lexies cibles,
- chercher le lien vers les prolexèmes de langue source puis vers les proaxies, ensuite vers les prolexèmes de langue cible, à la fin vers les lexies cibles,
- traiter l'étiquette,
- trier et afficher.

5 Résultats préliminaires

Nous avons séparé les trois niveaux de traduction pour afficher les résultats de recherche dans Jibiki: 1er – traduction directe par axie, 2ème – traduction par prolexème et proaxie avec la même étiquette, 3ème – traduction par prolexème et proaxie sans étiquette. Pour faciliter la lecture, nous avons décidé :

- d'afficher l'étiquette, la langue et le mot-vedette dans le 1er et le 2ème niveaux,
- d'afficher tous les détails (phrases d'exemple, définition, POS, etc.) dans le 3ème niveau, y compris les lexies du même prolexème de la langue source,
- de ne pas afficher la traduction dans le 2ème niveau si elle a déjà été trouvée et est déjà affichée dans le 1er niveau.

5.1 Scénario 1 : terme « UN » de l'anglais vers toutes les langues

Niveau	Lexies trouvés par théories	Lexies affichées par interface
--------	-----------------------------	--------------------------------

	Fra	Eng	Zho	Jpn	Fra	Eng	Zho	Jpn
1	ONU				ONU			
2	ONU			国連				国連
3	ONU, Nations unies, onusien, Organisation de nations unies		联合国	国際連合, 国連	ONU, Nations unies, onusien, Organisation de nations unies	UN, United Nations	联合国	国際連合, 国連

Table 3 - Trois niveaux de traduction et affichage : terme « UN » de l'anglais vers toutes les langues

UTILISATEUR :

ying profil utilisateur
se déloguer

CONSULTATION :

Mot :

UN

SOURCE :

anglais

Cible :

toutes les langues

Rechercher

Recherche avancée

Liste des dictionnaires

ARTICLES :

Créer

Éditer

RÉVISION :

Contributions

Ordre alphabétique

Tableau résumé

Exporter un volume

ADMINISTRATION :

Dictionnaires

Volumes

Entrées

Feuilles de style

Utilisateurs

Groupes

Caches de données

Éditer les news

Propriétés du serveur

Statistiques du serveur

Résultats de recherche

1 entrée(s) trouvées.

Traduction trouvé par la couche axie

Étiquette = ACRO Langue = fra **ONU**

Traduction trouvé par la couche proxie (avec la même étiquette)

Étiquette = ACRO Langue = jpn 国連

ProAxie Acro.proxie.United_Nations

ProAxeme Acro.prolex.eng.United_Nations.1

UN [n]

United Nations (Abbreviations UN)

FINISHED by edit duplicate & edit delete view history view XML

United Nations [n]

The United Nations (abbreviated UN in English, and ONU in French and Spanish), is an international organization whose stated aims are facilitating cooperation in international law, international security, economic development, social progress, human rights, and achievement of world peace. The UN was founded in 1945 after World War II to replace the League of Nations, to stop wars between countries, and to provide a platform for dialogue.

FINISHED by edit duplicate & edit delete view history view XML

ProAxeme Acro.prolex.fra.Organisation_des_nations_unies.1

ONU [n]

Initiales de Organisation des Nations Unies.

FINISHED by edit duplicate & edit delete view history view XML

Nations unies [n]

Alias d'Organisation des Nations Unies.

FINISHED by edit duplicate & edit delete view history view XML

Figure 5 : terme « UN » de l'anglais vers toutes les langues

UNUSATEUR :
 ying profil utilisateur
 se déloguer

CONSULTATION :
 Mot :
 UN
 Source :
 (anglais)
 Cible :
 toutes les langues

Recherche avancée
 Liste des dictionnaires

Ajouter :
 Créer
 Editer

Reviser :
 Contributions
 Ordre alphabétique
 Tableau résumé
 Exporter un volume

Administration :
 Dictionnaires
 Volumes
 Entrées
 Feuilles de style
 Utilisateurs
 Groupes
 Caches de données
 Editer les news
 Propriétés du serveur
 Statistiques du serveur

Résultats de recherche

1 entrée(s) trouvées.

Traduction trouvé par la couche axie
 Étiquette = ACRO Langue = fra **ONU**

Traduction trouvé par la couche proaxie (avec la même étiquette)
 Étiquette = ACRO Langue = jpn **国連**

ProAxie Acro.proaxie.United_Nations-

ProAxie Acro.prolex.eng.United_Nations.1

UN [n]
 United Nations (Abbreviations UN)

United Nations [n]
 The United Nations (abbreviated UN in English, and ONU in French and Spanish), is an international organization whose stated aims are facilitating cooperation in international law, international security, economic development, social progress, human rights, and achievement of world peace. The UN was founded in 1945 after World War II to replace the League of Nations, to stop wars between countries, and to provide a platform for dialogue.

ProAxie Acro.prolex.fra.Organisation_des_nations_unies.1

ONU [n]
 Initiales de Organisation des Nations Unies.

Nations unies [n]
 Alias d'Organisation des Nations Unies.

onusien [n]
 Alias d'Organisation des Nations Unies.

Organisation de nations unies [n]
 L'Organisation des Nations unies (ONU) est une organisation internationale regroupant, à quelques exceptions près, tous les États de la planète. Distincte des États qui la composent, l'organisation a pour finalité la paix internationale. Ses objectifs sont de faciliter la coopération dans le droit international, la sécurité internationale, le développement économique, le progrès social, les droits de l'homme et la réalisation à terme de la paix mondiale. L'ONU est fondée en 1945 après la Seconde Guerre mondiale pour remplacer la Société des Nations, afin d'arrêter les guerres entre pays et de fournir une plate-forme de dialogue.

ProAxie Acro.prolex.zho.联合国.1

联合国 [n]
 联合国（常以英文缩写UN表示）是一个由主权国家组成的国际组织，致力于促进各国在国际法、国际安全、经济发展、社会进步、人权及实现世界和平方面的合作。联合国成立于第二次世界大战结束后的1945年，用以取代国际联盟，去阻止战争并为各国提供对话平台。

ProAxie Acro.prolex.jpn.国際連合.1

国際連合 [n]
 国際連合（こくさいれんごう、英語: United Nations、略称は国連（こくれん）、UN）は、国際連合憲章の下、1945年に設立された国際組織である。主たる活動目的は国際平和の維持（安全保障）、そして経済や社会などに関する国際協力の実現である。なお、原語のUnited Nationsはもと第二次大戦中の枢軸国に対する連合国を指す言葉であり、中国語等では「国聯連合」ではなく「連合国」と呼ぶ。

国連 [n]
 国連連合（こくさいれんごう、英語: United Nations、略称は国連（こくれん）、UN）

Figure 6 : Affichage complet – terme « UN » de l'anglais vers toutes les langues

5.2 Scénario 2 : terme « 国連 » du japonais vers toutes les langues

Niveau	Lexies trouvés par théories				Lexies affichées par interface			
	Fra	Eng	Zho	Jpn	Fra	Eng	Zho	Jpn
1								
2	ONU	UN			ONU	UN		
3	ONU, Nations unies, Organisation de nations unies	UN, United Nations	联合国		ONU, Nations unies, Organisation de nations unies	UN, United Nations	联合国	国際連合, 国連

Table 4 - Trois niveaux de traduction et affichage : terme « 国連 » du japonais vers toutes les langues

Résultats de recherche

1 entrée(s) trouvées.

Traduction trouvé par la couche proaxie (avec la même étiquette)
Étiquette = ACRO Langue = fra **ONU**

Traduction trouvé par la couche proaxie (avec la même étiquette)
Étiquette = ACRO Langue = eng **UN**

ProAxie Acro.proaxie.United_Nations
ProAxeme Acro.prolex.eng.United_Nations.1

UN [n]
United Nations (Abbreviations UN)

United Nations [n]
The United Nations (abbreviated UN in English, and ONU in French and Spanish), is an international organization whose stated aims are facilitating cooperation in international law, international security, economic development, social progress, human rights, and achievement of world peace. The UN was founded in 1945 after World War II to replace the League of Nations, to stop wars between countries, and to provide a platform for dialogue.

ProAxeme Acro.prolex.fra.Organisation_des_nations_unies.1

ONU [n]
Initiales de Organisation des Nations Unies.

Nations unies [n]
Alias d'Organisation des Nations Unies.

Figure 7 : terme « 国連 » du japonais vers toutes les langues

5.3 Scénario 3 : terme « onusien » du français vers toutes les langues

Niveau	Lexies trouvés par théories				Lexies affichées par interface			
	Fra	Eng	Zho	Jpn	Fra	Eng	Zho	Jpn
1								
2								
3		UN, United Nations	联合国	国際連合, 国連	ONU, Nations unies, onusien, Organisation de nations unies	UN, United Nations	联合国	国際連合, 国連

Table 5 - Trois niveaux de traduction et affichage : terme « onusien » du français vers toutes les langues

The screenshot displays the 'Résultats de recherche' (Search Results) page of the ProAxie system. On the left, a sidebar contains navigation menus for 'UTILISATEUR', 'CONSULTATION', 'ARTICLES', 'RÉVISION', and 'ADMINISTRATION'. The main content area shows search results for the term 'onusien'. The first result is 'United Nations (Abbreviations UN)' with the label 'UN [n]'. The second result is 'United Nations [n]' with a detailed description: 'The United Nations (abbreviated UN in English, and ONU in French and Spanish), is an international organization whose stated aims are facilitating cooperation in international law, international security, economic development, social progress, human rights, and achievement of world peace. The UN was founded in 1945 after World War II to replace the League of Nations, to stop wars between countries, and to provide a platform for dialogue.' Below this, there are three more results: 'ONU [n]' (Initiales de Organisation des Nations Unies.), 'Nations unies [n]' (Alias d'Organisation des Nations Unies.), and 'onusien [n]' (Alias d'Organisation des Nations Unies.). Each result entry includes a 'FINISHED by' status and links for 'edit', 'duplicate & edit', 'delete', 'view history', and 'view XML'.

Figure 8 : terme « onusien » du français vers toutes les langues

6 Conclusion et perspectives

La gestion des terminologies avec liens riches est présentée avec un exemple d'acronyme (ONU) de nom propre dans cet article. Nous avons repris les concepts de lexie, d'axie, de prolexème, et établi le concept de proaxie pour produire la macrostructure de ProAxie. Dans cette macrostructure, une étiquette est utilisée pour représenter les relations de lexie. Nous avons implémenté la solution de la macrostructure ProAxie dans la plateforme Jibiki en utilisant la nouvelle extension Jibiki-Pivax, et créé trois niveaux de traduction en théorie et en affichage.

Pour le futur, nous souhaitons faire évoluer la macrostructure de ProAxie pour prendre en compte les synonymies de genre, et transposer le concept de prolexème pour que cette solution puisse être utilisée dans un autre domaine linguistique. Par exemple, pour une ressource lexicale comprenant des textos, en français « A+ » correspond à « À plus » avec une étiquette « texto », et en anglais « L8R » correspond à « later » avec l'étiquette « texto ».

Nous prévoyons de prendre en compte également les quatre variations du diasystème basé essentiellement sur ce que Eugenio Coseriu propose : diachronique (variété dans le temps), diaphasique (variété concernant les finalités de l'emploi), diatopique (variété dans l'espace) et diastratique (variété relative à la stratification socio-culturelle). Nous voudrions plus tard enrichir la notion d'étiquette.

Bibliography

[MT00] : Tran Mickael , ProlexbaseUn dictionnaire relationnel multilingue de noms propres : conception, implémentation et gestion en ligne, 2006, Thèse de doctorat d’informatique, pages 55-57

[MTMMEP00] : Tomokiyo Mutsuko and Mangeot Mathieu and Planas Emmanuel, Papillon: a Project of Lexical Database for English, French and Japanese, using Interlingual Links, 2001, In Actes de JST 2001, 3 p.

[GSMM01a] : Sérasset Gilles and Mangeot Mathieu, Papillon Lexical Database Project: Monolingual Dictionaries and Interlingual Links, 2001, In Proc. Of NLPRS 2001, pages 119-125

[MLCAPA00] : Mel’čuk Igor and Clas Andre and Polguère Alain, Introduction à la lexicologie explicative et combinatoire, 1995, Livre, 256 p.

[Francopoulo09] : Francopoulo Gil and Bel Nuria and George Monte and Calzolari Nicoletta and Monachini Monica and Pet Mandy and Soria Claudia, Multilingual resources for NLP in the lexical markup framework (LMF), 2009, In journal de Language Resources and Evaluation, March 2009, Volume 43, pages 57-70

[ThesisHTN] : Nguyen Hong-Thai, Des systèmes de TA homogènes aux systèmes de TAO hétérogènes, 2009, Thèse de doctorat d’informatique, 236 p.

[MMHTN09] : Mangeot Mathieu and Nguyen Hong-Thai, Building lexical resources: towards programmable contributive platforms, 2009, In Proc. IEEE-RIVF 2009, pages 84-92

[MM06] : Mangeot Mathieu, Dictionary Building with the Jibiki Platform, 2006, In Proc. of EURALEX 2006, 5 p.

[MM02] : Mangeot Mathieu , An XML Markup Language Framework for Lexical Databases Environments: the Dictionary Markup Language., 2002, In Proc of Papillon 2002 Workshop, 8 p.